Sentiment Analysis

A Probabilistic Approach

S. A. Gieske S. Laan D. S. Ten Velthuis
C. R. Verschoor A. J. Wiggers

Faculty of Science (FNWI)
University of Amsterdam

February 2, 2012

Outline

1 The goal

2 Approach

3 Data Preprocessing

4 Classification

5 Webserver Framework

6 Conclusion

Outline

1 The goal
2 Approach
3 Data Preprocessing
4 Classification
5 Webserver Framework
6 Conclusion

The goal of the project

Project Description

Performing sentiment analysis on messages about the EO

- Classification Sentiment vs. Non Sentiment
- Classification Positive vs. Negative

Outline

1 The goal

2 Approach

3 Data Preprocessing

4 Classification

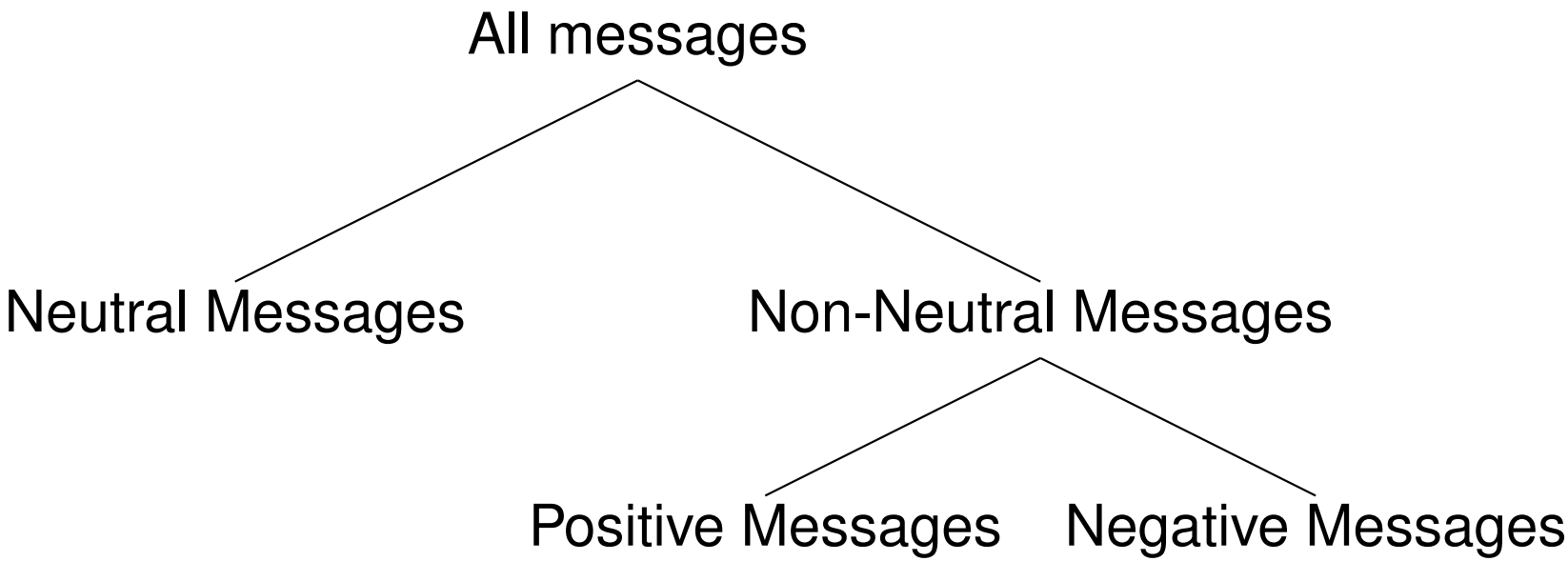
5 Webserver Framework

6 Conclusion

Approach

- Preprocessing of the data
- Perform machine learning algorithms on data
- Use best algorithms to classify real time on server

Hierarchical Classification



Outline

1 The goal

2 Approach

3 Data Preprocessing

4 Classification

5 Webserver Framework

6 Conclusion

Outline

3 Data Preprocessing
- Dataset Analysis
- Data Cleaning
- Data Reduction
- Data Manipulation

Dataset Analysis

Dataset messages EO

10.000 messages, 19 features per message

Ony 3 features used:

- Source
- Sentiment
- Message contents

Data Cleaning

- Shorten words, ex. hahaha to haha
- Stemmer

Data Reduction

- Only use Twitter messages (83% of all messages)
- Remove articles, reference words and prepositions
- Substitute smileys with words
- Remove some punctuation marks (ex. not ! ?)

Data Manipulation

- Create tokens
- Assign sentiment probabilities to tokens

$$P(word) = \frac{\sum word \in C_1}{\sum word \in C_1 \cup C_2} \quad (1)$$

Outline

1 The goal

2 Approach

3 Data Preprocessing

4 Classification

5 Webserver Framework

6 Conclusion

Outline

4 Classification
- Perceptron
- Support Vector Machine
- Naive Bayes
- Multiclassification with Perceptron
- Entropy
- Neural Network

Classification

Perceptron

Results

Support Vector Machine

Results

Naive Bayes

Results

Multiclassification with Perceptron

Results

Entropy

Results

Neural Network

Results

Outline

1 The goal

2 Approach

3 Data Preprocessing

4 Classification

5 Webserver Framework

6 Conclusion

Webserver Framework

Request (HTML) $\rightarrow$ Server (PHP/PYTHON) $\rightarrow$ Result (XML)

Request http://url.com/?**dataset**=1&**message**=De EO is cool!

Result XML File (Containing: Status, Message, Sentiment, Accuracy, Precision, Recall)

Demo

Action...

Outline

1 The goal

2 Approach

3 Data Preprocessing

4 Classification

5 Webserver Framework

6 Conclusion

Conclusion

Conclusion